I spent more than two hours on this, but I probably overdid the analysis.

We begin by giving an analysis in terms of the mean and standard deviation of the observed female performance for gun time, net time, and pace. This analysis assumes the data are distributed approximately normally. As a test of the validity of the assumption of normality we include the Shapiro-Wilk (SW) statistic and its associated p-value. The observed values are given in Table 1.

Table 1 – Analysis in Terms of Mean and Standard Deviation for Females

|  |  |  |  |
| --- | --- | --- | --- |
|  | Gun Time | Net Time | Pace |
| Mean | 62 min | 58 min | 9.4 min/mile |
| Standard Dev | 11 min | 9.9 min | 1.6 min/mile |
| SW Statistic | 0.97 | 0.95 | 0.95 |
| SW P-value | 2.8e-14 | 1.1e-18 | 1.1e-18 |

We find that the mean net time is 4 minutes lower than the mean gun time. We expected this to be lower by definition. It appears a participant averages about 4 minutes to cross the starting line after the gun.

We observe that the very small values of the Shapiro-Wilk p-value are consistent with a rejection of the assumption of normality for these distributions. This is good, since we should not expect these values to be normally distributed since there is a definite lower bound in each case while there is no upper bound and we expect a long tails for longer times and slower paces.

We repeat this analysis for males. The results are given in Table 2.

Table 2 – Analysis in Terms of Mean and Standard Deviation for Males

|  |  |  |  |
| --- | --- | --- | --- |
|  | Gun Time | Net Time | Pace |
| Mean | 55 min | 52 min | 8.4 min/mile |
| Standard Dev | 11 min | 9.7 min | 1.6 min/mile |
| SW Statistic | 0.98 | 0.97 | 0.97 |
| SW P-value | 7.8e-12 | 8.6e-16 | 8.7e-16 |

We find that the mean net time is 3 minutes lower than the mean gun time. Also the mean gun time is 7 minutes lower for males than for females, the mean net time is 6 minutes lower, and the mean pace is 1 min/mile lower.

Again, the Shapiro-Wilk statistic is consistent with the distributions not being normal.

The non-normality of the data indicates we should probably proceed with an analysis in terms of the median and associated quantiles.

We next provide an analysis in terms of the median and interquartile range (IQR) of the observed female performance for gun time, net time, and pace. The observed values are given in Table 3.

Table 3 – Analysis in Terms of Median and Interquartile Range (IQR) for Females

|  |  |  |  |
| --- | --- | --- | --- |
|  | Gun Time | Net Time | Pace |
| Median | 61 min | 58 min | 9.3 min/mile |
| IQR | 13 min | 11 min | 1.8 min/mile |

We observe that the calculated medians do not differ by much from the means calculated previously. Also, the median net time is 3 minutes lower than the median gun time.

We repeat the analysis for males. The observed values are given in Table 4.

Table 4 – Analysis in Terms of Median and Interquartile Range (IQR) for Males

|  |  |  |  |
| --- | --- | --- | --- |
|  | Gun Time | Net Time | Pace |
| Median | 54 min | 51 min | 8.3 min/mile |
| IQR | 14 min | 12 min | 2.0 min/mile |

Again, the calculated medians do not differ much from the previously calculated means. We observe that male median gun time is 7 minutes less than the female one, the male median net time is also 7 minutes less, and the pace is 1.0 min/mile faster.

We can obtain some more information about how the gum times, net times, and paces compare between females and males by looking at the associated box plots. These are given in Figures 1 and 2.

Figure 1 - Box Plots of Gun Times and Net Time for Females (F) and Males (M)

A diagram of a number of objects

AI-generated content may be incorrect.

Figure 2 – Box Plots of Paces for Females (F) and Males (M)

A diagram of a graph

AI-generated content may be incorrect.

The whiskers in these plots are given by 1.5 times the IQR, and points outside of the whiskers are considered to be outliers. The existence of outlier values for participants having very slow times and paces is what we would expect. It is interesting that the fastest female times and paces are outliers in the female distributions while the same is not true for the males.

We next give the analysis in terms of mode for females, Table 5, and males, Table 6.

Table 5 – Analysis in Terms of Mode for Females

|  |  |  |  |
| --- | --- | --- | --- |
|  | Gun Time | Net Time | Pace |
| Mode | 62 min | 48 min | 9.0 min/mile |

Table 6 – Analysis in Terms of Mode for Males

|  |  |  |  |
| --- | --- | --- | --- |
|  | Gun Time | Net Time | Pace |
| Mode | 55 min | 35 min | 7.3 min/mile |

The modes for the gun times are close to the means and medians for both females and males. The modes for net times are 10 min and 16 min lower than the medians for females and males, respectively. The modes of the pace is close to the median for females, while that for males is 1.0 min/mile lower. These results are consistent with a larger group of people coming in together with faster times but with a long tail of slower people coming in afterwards over time.

The gun time gives the time for a participant to finish the race from the time of firing the gun while net time gives the time from the moment the participant crossed the starting line. We can analyze the time period between the firing of the gun and when the participant crossed the starting line. We provide an analysis in terms of mean and standard deviation for females, Table 7, and males, Table 8.

Table 7 – Analysis in Terms of Mean and Standard Deviation for Gun Times – Net Times for Females

|  |  |
| --- | --- |
|  | Gun Times - Net Times |
| Mean | 3.3 min |
| Standard Dev | 1.5 min |
| SW Statistic | 0.92 |
| SW P-value | 5.1e-24 |

Table 8 – Analysis in Terms of Mean and Standard Deviation for Gun Times – Net Times for Males

|  |  |
| --- | --- |
|  | Gun Times - Net Times |
| Mean | 2.5 min |
| Standard Dev | 1.6 min |
| SW Statistic | 0.92 |
| SW P-value | 2.2e-25 |

We see that the mean value for males is 0.8 min lower than that for females. Looking at the SW p-value we see that the distributions for gun times – net times fail the test of normality in both cases.

Table 9 – Analysis in Terms of Median and Interquartile Range (IQR) for Gun Times – Net Times for Females

|  |  |
| --- | --- |
|  | Gun Times - Net Times |
| Mean | 3.5 min |
| IQR | 0.97 min |

Table 10 – Analysis in Terms of Median and Interquartile Range (IQR) for Gun Times – Net Times for Males

|  |  |
| --- | --- |
|  | Gun Times - Net Times |
| Mean | 3.1 min |
| IQR | 2.5 min |

We see that the median value for the males is only 0.4 min lower than that for females.

We compare the corresponding box plots in Figure 3.

Figure 3 – Box Plots of Gun Times Minus Net Times for Females (F) and Males (M)

A graph of an object and object

AI-generated content may be incorrect.

We note that the central 50% of the female times are very tight while the tails are very heavy with a large number of outliers at both the short and long time ends. The male times are more spread out so that very few participants qualify as outliers.

It is interesting to consider whether the time taken from gun to starting line has any relationship with the overall race performance. We therefore calculate the correlations between gun times – net times and gun times and net times using Pearson’s R statistic. The results are given in Tables 11 and 12 for females and males, respectively.

Table 11 – Correlations for Gun Times – Net Times for Females

|  |  |  |
| --- | --- | --- |
|  | Gun Times | Net Times |
| R | 0.70 | 0.62 |
| p-value | 1.3e-163 | 2.7e-118 |

Table 12 – Correlations for Gun Times – Net Times for Males

|  |  |  |
| --- | --- | --- |
|  | Gun Times | Net Times |
| R | 0.77 | 0.70 |
| p-value | 4.7e-246 | 1.2e-183 |

We observe that both females and males exhibit positive correlations in both cases. So higher gun times – net times correlate to higher gun times and higher net times. That is faster runners overall try to get past the starting line faster. The correlation is stronger for males in both cases.

We can look at this behavior in the set of scatter plots given in Figure 4.

Figure 4 – Scatter Plots of Gun Times and Net Times Versus Gun Times Minus Net Times for Females (F) and Males (M)

A group of blue dots

AI-generated content may be incorrect.

The correlation does appear to be present in the plots and not due to outliers. I am not sure what the consistent grouping of the data into 4 main groups means. I would need more information about how this kind of race is conducted.

At this point we must divide the participant data into divisions. Divisions are given by the combination of age groups and genders. The PDF seems to say that the age groups are 0-14, 15-19, 20-29, 30-39 and so on. The data value Div/Tot given in the data seems like it might be corresponding to divisions of 0-14, 15-19, 20-24, 25-29, 30-34 and so on. I say this because the male age range of 40-49 looked like it consisted of 2 Div/Tot groups which corresponded to 40-44 and 45-49. I am proceeding by dividing the data into age groups corresponding to how I understand the PDF.

At this point we look specifically at Chris Doe. Chris Doe falls into the male age 40-49 division. We compare his performance to the 90th percentile of this division. The 90th percentile performances for his division were 43 min for gun time, 42 min for net time, and 6.7 min/mile for pace. We find that he ran a 10 min slower gun time, an 8 min slower net time, and a 1.3 min/mile lower pace than the 90th percentile of his division.

We finally look at the performance in the race by division. We restrict to net times to factor out the issue of the time taken to cross the starting line after the gun. We begin with females. Figure 5 gives the box plots for all divisions for females. Figure 6 gives the numbers of participants for all divisions for females.

Figure 5 – Box Plots of Net Times by Age Group for Females

A graph with numbers and lines

AI-generated content may be incorrect.

Figure 6 – Plot of Numbers of Participants in Each Age Group for Females

A graph with numbers and a triangle

AI-generated content may be incorrect.

We first observe that the numbers of participants for divisions 0-14, 15-19, 60-69, and 70-79 are significantly lower than for the other divisions. It is likely that these groups represent a less broad sample of the age range than do the other divisions and likely are weighted to higher performers in that age range. The medians and 2nd and 3rd quartiles for 20-29, 30-39, and 40-49 are about the same, while the median and 2nd and 3rd quartiles for 50-59 have shifted somewhat to values indicating slower running speeds. The 20-29, 30-39, and 40-49 divisions all exhibit long tails for slower times as indicated by the outliers on this end. This indicates higher participation by less competitive runners in these age ranges. The top performers for 20-29, 30-39, and 40-49 are outliers. Again, the seemingly better performance for ages 60-69 and 70-79 probably indicates participation by only higher performers. The same is probably the case for ages 15-19.

We continue with males. Figure 7 gives the box plots for all divisions for males. Figure 8 gives the numbers of participants for all divisions for males.

Figure 7 - Box Plots of Net Times by Age Group for Males

A graph with numbers and lines

AI-generated content may be incorrect.

Figure 8 – Plot of Numbers of Participants in Each Age Group for Males

A graph with numbers and lines

AI-generated content may be incorrect.

Again, we observe that the numbers of participants for divisions 0-14, 15-19, 60-69, 70-79, and 80-89 are significantly lower than for the other divisions. The remaining divisions show an increase in the median net time as well as the 2nd and 3rd quartiles with age. The age groups 20-29, 30-39, and 40-49 show participation by a larger number of less competitive individuals as indicated by the outliers at high net times, although the other ages show more outliers than was the case with females. In no division was the highest performing male an outlier.